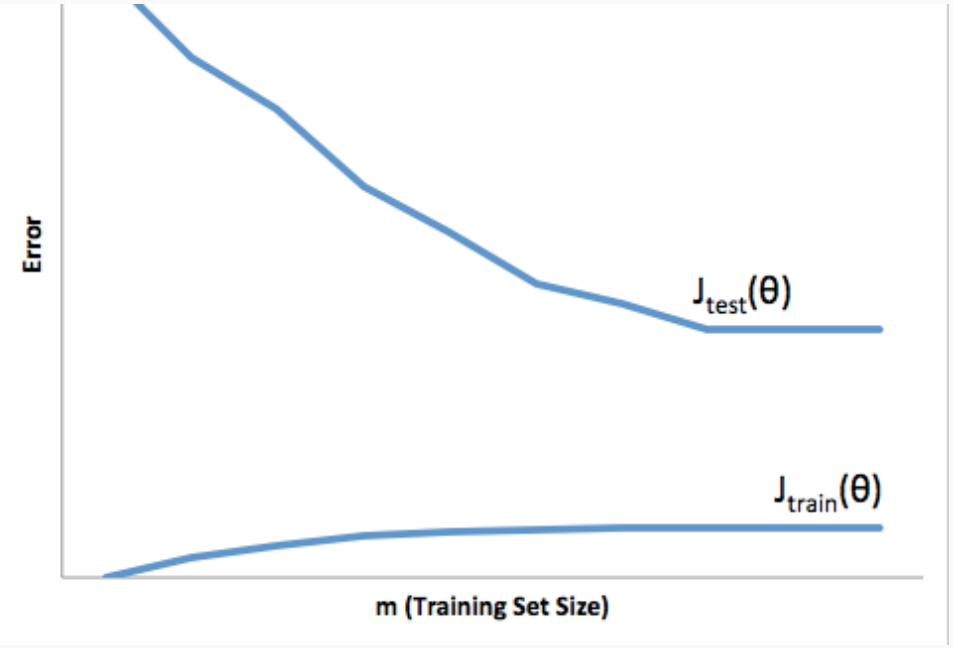
Feedback — X. Advice for Applying Machine Learning

Help

You submitted this quiz on **Sun 27 Apr 2014 6:14 AM IST**. You got a score of **4.75** out of **5.00**. You can attempt again in 10 minutes.

Question 1

You train a learning algorithm, and find that it has unacceptably high error on the test set. You plot the learning curve, and obtain the figure below. Is the algorithm suffering from high bias, high variance, or neither?



Your Answer		Score	Explanation
○ Neither			
◉ High variance	✔	1.00	This learning curve shows a large gap between training and test set errors, so the algorithm is suffering from high variance.
○ High bias			
Total		1.00 / 1.00	

Question 2

Suppose you have implemented regularized logistic regression to classify what object is in an image (i.e., to do object recognition). However, when you test your hypothesis on a new set of images, you find that it makes unacceptably large errors with its predictions on the new images. However, your hypothesis performs **well** (has low error) on the training set. Which of the following are promising steps to take? Check all that apply.

Your Answer		Score	Explanation
☑ Try decreasing the regularization parameter λ.	✖	0.00	The gap in errors between training and test suggests a high variance problem in which the algorithm has overfit the training set. Decreasing the regularization parameter will increase the overfitting, not decrease it.
☑ Try using a smaller set of features.	✔	0.25	The gap in errors between training and test suggests a high variance problem in which the algorithm has overfit the training set. Reducing the feature set will ameliorate the overfitting and help with the variance problem.
☐ Try adding polynomial features.	✔	0.25	The gap in errors between training and test suggests a high variance problem in which the algorithm has overfit the training set. Using more complex features will only increase the overfitting of the training set.
☑ Get more training examples.	✔	0.25	The gap in errors between training and test suggests a high variance problem in which the algorithm has overfit the training set. Adding more training data will increase the complexity of the training set and help with the variance problem.
Total		0.75 / 1.00	

Question 3

Suppose you have implemented regularized logistic regression to predict what items customers will purchase on a web shopping site. However, when you test your hypothesis on a new set of customers, you find that it makes unacceptably large errors in its predictions. Furthermore, the hypothesis performs **poorly** on the training set. Which of the following might be promising steps to take? Check all that apply.

Your Answer		Score	Explanation
☑ Try to obtain and use additional features.	✔	0.25	The poor performance on both the training and test sets suggests a high bias problem. Using additional features will increase the complexity of the hypothesis, thereby improving the fit to both the train and test data.
☐ Use fewer training examples.	✔	0.25	Using fewer training examples should never improve test set performance, as the model has fewer data points from which to learn.
☐ Try using a smaller set of features.	✔	0.25	The poor performance on both the training and test sets suggests a high bias problem. Using fewer features will decrease the complexity of the hypothesis and will make the bias problem worse.
☑ Try adding polynomial features.	✔	0.25	The poor performance on both the training and test sets suggests a high bias problem. Adding more complex features will increase the complexity of the hypothesis, thereby improving the fit to both the train and test data.
Total		1.00 / 1.00	

Question 4

Which of the following statements are true? Check all that apply.

Your Answer		Score	Explanation
☐ Suppose you are training a regularized linear regression model.The recommended way to choose what value of regularization parameter λ to use is to choose the value of λ which gives the lowest **training set** error.	✔	0.25	You should not use training error to choose the regularization parameter, as you can always improve training error by using less regularization (a smaller value of λ). But too small of a value will not generalize well on the test set.
☑ Suppose you are training a regularized	✔	0.25	The cross validation lets us find the "just right" setting of the regularization parameter given the fixed model

linear regression model. The recommended way to choose what value of regularization parameter λ to use is to choose the value of λ which gives the lowest **cross validation** error.			parameters learned from the training set.
Suppose you are training a regularized linear regression model. The recommended way to choose what value of regularization parameter λ to use is to choose the value of λ which gives the lowest **test set** error.	✔	0.25	You should not use the test set to choose the regularization parameter, as you will then have an artificially low value for test error and it will not give a good estimate of generalization error.
A typical split of a dataset into training, validation and test sets might be 60% training set, 20% validation set, and 20% test set.	✔	0.25	This is a good split of the data, as it dedicates the bulk of the data to finding model parameters in training while leaving enough data for cross validation and estimating generalization error.
Total		1.00 / 1.00	

Question 5

Which of the following statements are true? Check all that apply.

Your Answer		Score	Explanation
If the training and test errors are about the same, adding more features will **not** help improve the results.	✔	0.25	If the two errors are the same, then the model has high bias, so adding more features will be helpful.

When debugging learning algorithms, it is useful to plot a learning curve to understand if there is a high bias or high variance problem.	✔	0.25	The shape of a learning curve is a good indicator of bias or variance problems with your learning algorithm.
A model with more parameters is more prone to overfitting and typically has higher variance.	✔	0.25	More model parameters increases the model's complexity, so it can more tightly fit data in training, increasing the chances of overfitting.
If a learning algorithm is suffering from high variance, adding more training examples is likely to improve the test error.	✔	0.25	With high variance, the model is overfitting the training data. Adding more training data will increase the complexity of the the train set, thereby reducing the chances of overfitting.
Total		1.00 / 1.00	